Three-dimensional structure of the AAH26994.1 protein from *Mus musculus*, a putative eukaryotic Urm1

SHANTERI SINGH, MARCO TONELLI, ROBERT C. TYLER, ARASH BAHRAMI, MIN S. LEE, AND JOHN L. MARKLEY

Center for Eukaryotic Structural Genomics, Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706-1544, USA

(RECEIVED May 8, 2005; FINAL REVISION May 8, 2005; ACCEPTED May 11, 2005)

Abstract

We have used NMR spectroscopy to determine the solution structure of protein AAH26994.1 from *Mus musculus* and propose that it represents the first three-dimensional structure of a ubiquitin-related modifier 1 (Urm1) protein. Amino acid sequence comparisons indicate that AAH26994.1 belongs to the Urm1 family of ubiquitin-like modifier proteins. The best characterized member of this family has been shown to be involved in nutrient sensing, invasive growth, and budding in yeast. Proteins in this family have only a weak sequence similarity to ubiquitin, and the structure of AAH26994.1 showed a much closer resemblance to MoaD subunits of molybdopterin synthases (known structures are of three bacterial MoaD proteins with 14%–26% sequence identity to AAH26994.1). The structures of AAH26994.1 and the MoaD proteins each contain the signature ubiquitin secondary structure fold, but all differ from ubiquitin largely in regions outside of this fold. This structural similarity bolsters the hypothesis that ubiquitin and ubiquitin-related proteins evolved from a protein-based sulfide donor system of the molybdopterin synthase type.

Keywords: Urm1; NMR spectroscopy; structural genomics; ubiquitin fold

Ubiquitin (Ub) and ubiquitin-like modifier proteins (Ubls) are small proteins that act primarily by conjugation to target proteins. Most Ubls have a C-terminal Gly–Gly, which is activated by adenylation and then reacts with the thiol group of a Cys residue on the activating enzyme (E1) to form a thioester linkage (E1–Ubl). The activated Ubl can be transferred to a Cys residue on a conjugating enzyme (E2) and next to the Lys residue of a target protein as catalyzed by a ligase (E3) (Hershko et al. 2000; Hochstrasser 2000). This process is highly regulated in cells and involves the sequential activity of the activating (E1), conjugating (E2), and ligating (E3) enzymes.

Reprint requests to: John L. Markley, Department of Biochemistry, University of Wisconsin–Madison, 433 Babcock Drive, Madison, WI 53706, USA; e-mail: markley@nmrfam.wisc.edu; fax: (608) 262-3179.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.051577605.

The Center for Eukaryotic Structural Genomics (CESG) chose *Mus musculus* AAH26994.1 as a target on the basis of its low-sequence identity with proteins of known structure in the Protein Data Bank (PDB). The sequence of this protein suggested that it is a member of the recently described ubiquitin-related modifier protein (Urm1) family (Furukawa et al. 2000). Furukawa and coworkers characterized Urm1, along with its unique E1-like activating protein (Uba4) as the fifth conjugation system in yeast, along with ubiquitin, Smt3, Rub1, and Apg12. They also noted the sequence resemblance of these proteins to proteins from prokaryotes and higher organisms involved in sulfur transfer. Specifically, Urm1 showed an ~40% sequence identity to *Escherichia coli* modifier-like proteins from the molybdopeterin (MoaD) and thiamin (ThiS) biosynthetic pathways, and Uba4 showed sequence similarity to the corresponding activating enzymes (MoeB and ThiF).

```
Mus musculus        AAPLCVKVEFGGGAELLFDGVKKHQVALPGQEEPWDIRNLLVWIKKNLLKERPELFIQGD
R. norvegicus       AAPLCVEVEFGGGAELLFDGVKKHQVTLPGQEEPWDIRNLLVWIKTNLLKERPELFIQGD
H. sapiens          AAPLSVEVEFGGGAELLFDGIKKHRVTLPGQEEPWDIRNLLIWIKKNLLKERPELFIQGD
G. gallus           AAPVSLQVEFGGGAELLFDGVKKHQVTLPSQPEPWDIRNLLKWIKQNLLKERPELFMQGE
D. melanogaster     ---LKIILEFSAGAELLFGNIKRRELNLDGKQK-WTIANLLKWMHANILTERPELFLQGD
A. thaliana         ---LEFGGGLELLCDSEKIHKVNVDLPNGADSDDFTMKHLLSWVRTNLIKERPEMFMKGD
S. cerevisiae       MVNVKVEFLGGLDAIFGKQRVHKIKMD-KEDPVTVGDLIDHIVSTMINNPNDVSIFIEDD

Mus musculus        SVRPGILVLINDADWELLGELDYQLQDQDSILFISTLHGG
R. norvegicus       SVRPGILVLINDADWELLGELDYQLQDQDSILFISTLHGG Identity 97/100(97%)
H. sapiens          SVRPGILVLINDADWELLGELDYQLQDQDSVLFISTLHGG Identity 93/100(93%)
G. gallus           SVRPGILVLINDADWELMGELDYKLQDQDNVLFISTLHGG Identity 85/100(85%)
D. melanogaster     TVRPGILVLINDTDWELLGELDYELQPNDNVLFISTLHGG Identity 62/97(63%)
A. thaliana         TVRPGVLVLVNDCDWELSGQLDTVIEDKDVVVFISTLHGG Identity 53/95(55%)
S. cerevisiae       SIRPGIITLINDTDWELEGEKDYILEDGDIISFTSTLHGG Identity 44/100(44%)
```

Figure 1. Sequence alignment of putative Urm1 from different organisms.

Urm1 was found to form a thioester with Uba4 (Furukawa et al. 2000); however, its natural E2, E3, and target substrate are unknown. As with ubiquitin and other ubiquitin-related modifiers, Urm1 possesses a Gly–Gly motif at the C terminus. Urm1 is found in many eukaryotes, including human (Fig. 1). An InterProScan search of the AAH26994.1 sequence showed that the protein shares 14%–22% sequence identity with the MoaD class of proteins, but only 6%–14% sequence identity with other ubiquitin-like proteins.

The NMR solution structure of the protein AAH26994.1 from *Mus musculus*, which we report here, reveals that AAH26994.1 possesses a β-grasp superfamily fold similar to those of MoaD, ThiS, and ubiquitin. Of the structures in the PDB, AAH26994.1 most closely resembles those of two prokaryotic MoaD proteins.

Results and Discussion

Resonance assignment and secondary structure of AAH26994.1

AAH26994 is an 11.5-kDa protein that consists of 101 residues, including five prolines. The ^{1}H-^{15}N HSQC spectrum of AAH26994.1 (Fig. 2) under the solution conditions used (1 mM protein concentration, 10 mM Bis-Tris buffer with 50 mM NaCl [pH 6.5], 25°C) revealed peaks with uniform shape and intensity indicative of a well-folded protein with 94 of the expected 95 backbone amide cross-peaks were observed in the spectrum. Amide backbone ^{15}N T_2 relaxation measurements (data not shown), which yielded an average T_2 value of 100 ± 10 msec, indicated that AAH26994.1 is monomeric under these conditions. We used a novel automated assignment software package, PISTACHIO (Eghbalnia et al. 2005; available from http://bija.nmrfam.wisc.edu/PISTACHIO), to determine backbone and side-chain assignments. To check these values, we carried out manual assignments. The final manual assignments used additional data from (H)CCH TOCSY to verify or correct the automated assignments and to include missing resonance assignments (see Materials and Methods). PISTACHIO utilized three-dimensional HNCO, HNCACB, CBCA(CO)NH, HBHA(CO)NH, H(CCO)NH, and C(CO)NH data sets as input and returned correct assignments for 97% of the backbone and 85% of the side-chain residues.

Elements of secondary structure were determined from secondary chemical shifts (Wishart and Sykes 1994) and through assignment and analysis of ^{15}N-edited and ^{13}C-edited NOESY-HSQC spectra. Figure 3 shows the summary of NMR parameters determined for

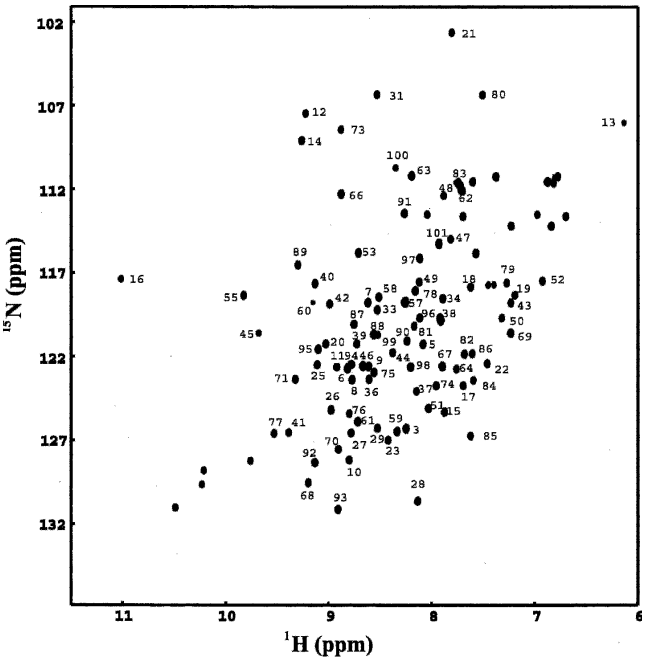

Figure 2. 2D ^{15}N-^{1}H HSQC spectrum of protein AAH26994.1. Resonance assignments are indicated by the residue number. Unnumbered peaks correspond to the side-chain NH of Trp or the amide side-chain groups of Asn or Gln.

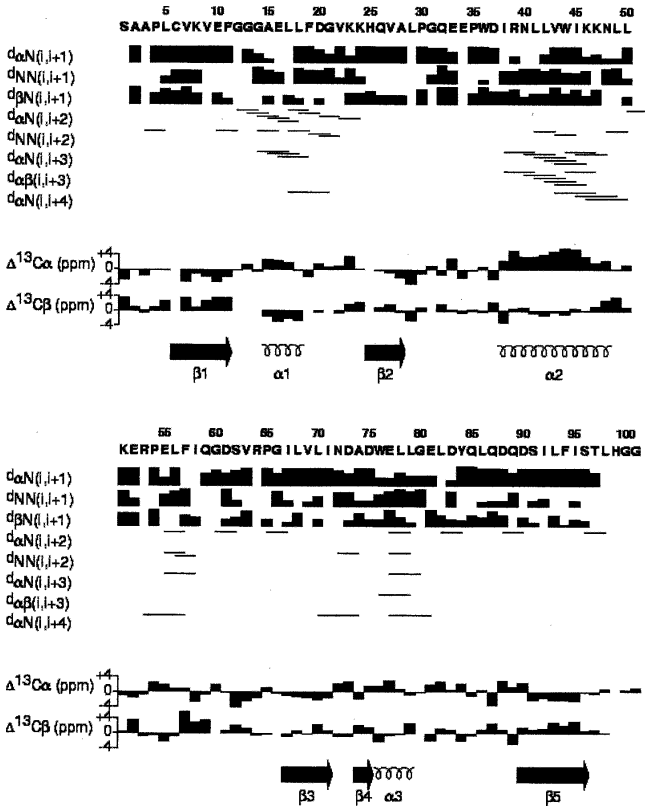

Figure 3. Amino acid sequence of AAH26994.1 and a summary of associated $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ secondary shifts and local backbone NOE connectivities.

AAH26994.1, which identified the following elements of secondary structure: five β-strands, two α-helices, and one 3_{10}-helix. Analysis of the ^{15}N-edited and ^{13}C-edited NOESY spectra revealed that the five β-strands form an anti-parallel β-sheet. The program TALOS (Cornilescu et al. 1999), which matches chemical shift and primary sequence with a database to predict ψ and ϕ angles, also corroborated these secondary structural elements within AAH26994.1.

Three-dimensional structure

The program ARIA (Nilges and O'Donoghue 1998) was used to calculate an ensemble of 100 conformers AAH26994.1 by simulated annealing on the basis of 1937 unambiguous and 229 ambiguous distance restraints. Additional restraints used for final structure calculation consisted of 96 dihedral angle constraints obtained from TALOS on the basis of backbone chemical-shift values and 46 hydrogen-bond restraints. The 20 conformers with the lowest total energy were used for final analysis. The global fold was established by using unambiguously assigned long-range NOEs. Several long-range NOEs between the helices were obtained from a 3D-NOESY(^{13}C,^{1}H-HSQC spectrum and a 3D-NOESY(^{15}N,^{1}H)-HSQC spectrum. Nearly 70% of the NOEs in the 3D-NOESY(^{15}N,^{1}H)-HSQC and nearly 30% of the NOEs in the 3D-NOESY(^{13}C,^{1}H)-HSQC were assigned manually. The unassigned NOEs with multiple possible assignments were treated by ARIA as ambiguous restraints. A summary of the experimental restraints and structural statistics is provided in Table 1. A stereo view of superposition of the final ensemble of 20 structures is shown in Figure 4A, and minimized-average structure depicted as ribbons is shown in Figure 4B. The 20 best structures showed a backbone RMSD of 0.55 ± 0.11 Å for the backbone atoms (Fig. 4C) and 1.22 ± 0.19 Å for all heavy atoms (for residues 4–98). The ensemble of 20 structures was analyzed by the PROCHECK-NMR (Laskowski et al. 1996) package (Table 1). Secondary structural elements included two

Table 1. *Structure statistics for AAH26994.1*

RMSD (Å) with respect to mean:	
Heavy backbone atoms (residues 4–98)	0.55 ± 0.11
All heavy atoms (residues 4–98)	1.22 ± 0.19
No. of experimental restraints	
Intra-residue NOEs	1039
Inter-residue sequential NOEs ($\|i-j\| = 1$)	296
Inter-residue medium range NOEs ($1 < \|i-j\| < 5$)	655
Inter-residue long range NOEs ($\|i-j\| > 4$)	229
Total NOEs	2219
Dihedral angle restraints	96
H-bond restraints	46
Restraint violations[a]	
NOE distances with violations >0.3 Å	3.2 ± 0.9
Dihedrals with violations >3°	1.07 ± 0.17
RMSD for experimental restraints[b]	
All distance restraints (2315) (Å)	0.037 ± 0.008
Torsion angles (96) (°)	1.07 ± 0.17
CNS energies from SA[c]	
Fvdw (kcal mol^{-1})	-950 ± 32
Felec (kcal mol^{-1})[d]	-3610 ± 102
RMSD (Å) from idealized covalent geometry	
Bonds (°)	0.0045 ± 0.00
Angles (°)	0.62 ± 0.02
Impropers (°)	1.76 ± 0.07
Ramachandran analysis (residues 1–101)	
Residues in the favored region (%)	75.0
Residues in additional allowed regions (%)	18.5
Residues in generously allowed regions (%)	3.72
Residues in disallowed regions (%)	2.8

Based on the 20 best structures and obtained by simulated annealing in CNS followed by refinement in explicit water using NOE distance restraints, dihedral angle restraints, bonds, angles, impropers, dihedral angle, van der Waals, and electrostatic energy terms.

[a] No distances were violated by >0.5 Å, and no dihedral angle restraints were violated by >5°.

[b] The number of each class of experimental restraints is given in parentheses.

[c] Force constants were described in Materials and Methods.

[d] The Lennard-Jones 6–12 and coulomb energy terms were calculated within CNS using the OPLS nonbonded parameters (as described in Materials and Methods).

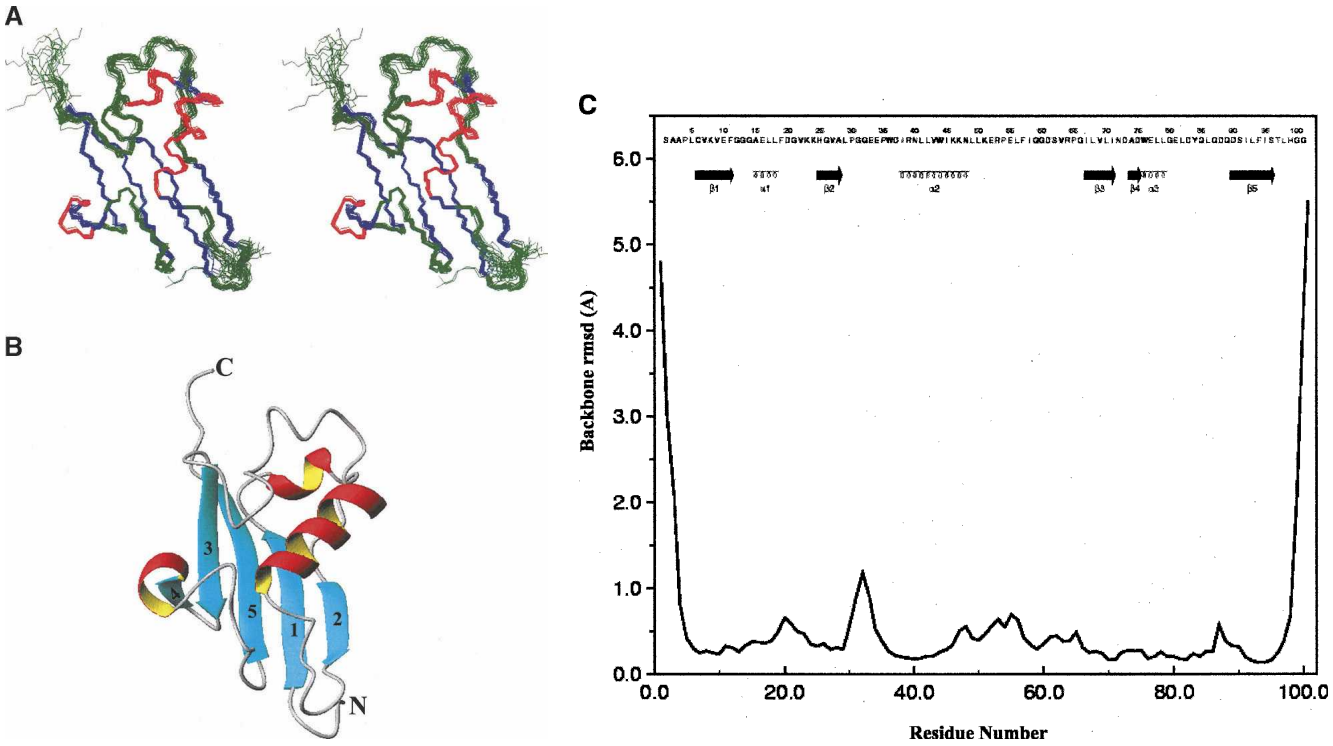

Figure 4. NMR solution structure of AAH26994.1. (*A*) Stereo view of the ensemble of 20 conformers that represent structure (backbone of residues 1–101). Coloring scheme: β-strand residues, blue; α-helical residues, red; other residues, green. (*B*) Ribbon view of the representative AAH26994.1 structure (that closest to the average) showing residues 1–101. Numbers represent the order of β-strands, and the N and C termini of the protein are indicated. The molecular graphics program MOLMOL (Koradi et al. 1996) was used in generating these views of the structure. (*C*) Backbone (N, C^{α}, C′) atomic RMSD plotted as a function of residue number.

α-helices (residues 38–48 and 76–79), one 3_{10}-helix (residues 15–18), and five β-strands (Fig. 4B). Strands 1 (residues 6–11) and 5 (residues 90–96) were found to run parallel to one another, while strands 2 (residues 25–28), 3 (residues 67–71), and 4 (residues 74–75) formed an anti-parallel conformation. This type of mixed β-sheet is characteristic of the compact ubiquitin fold. The hydrophobic core, which stabilizes the tertiary structure of AAH26994.1, is formed mainly by the concave surface imposed by the hydrophobic residues of β-sheets 1, 2, 3, and 5. Hydrophobic residues located in helix 1 and the 3_{10}-helix also support formation of the hydrophobic core, but to a lesser extent. The aromatic rings of Tyr36 and Tyr86, which are partially exposed, cause upfield ring-current shifts of proton resonances in residues Leu29, Val69, and Glu81; this is a phenomenon observed in ubiquitin type folds.

Fold classification and putative function

The classification of these proteins comes broadly under the Urm1, MoaD, and ThiS family. ThiS is a sulfur carrier protein that plays a central role in thiamine biosynthesis in *E. coli* (Wang et al. 2001). Although the relationships among Urm1, ThiS, and MoaD are firmly established by sequence homology, none of the proteins share significant sequence identity with ubiquitin. Despite the overall structural similarity of AAH26994.1, the degree of sequence identity is very low (7%). The structure of AAH26994.1 differs from that of ubiquitin (Ub) largely in the regions outside of the well-defined Ub signature secondary structure fold. However, the C-terminal Gly–Gly motif present in this protein also occurs in ubiquitin.

The minimum energy structure of AAH26994.1 was submitted to the fold recognition programs DALI (Holm and Sander 1995) and VAST (Gibrat and Madej 1996). Both programs identified the fold as ubiquitin type. The structure of AAH26994.1was closer to those of three bacterial MoaD proteins (subunits of molybdopterin synthase) than to that of any eukaryotic ubiquitin. The C^{α} RMSD between AAH26994.1 and each structural homolog, 1VJK (from *Pyrococcus furiosus*), 1FMA (from *E. coli*), and 1V8C (from *Thermus thermophilus* HB8), was calculated to be 1.9 Å, 2.0 Å, and 1.5 Å, respectively, with corresponding *Z*-scores of 6.3, 6.6, and 6.5. The sequence identities between AAH26994.1 and those in the 1VJK, 1FMA, and 1V8C structures are 26%, 23%, and 14%,

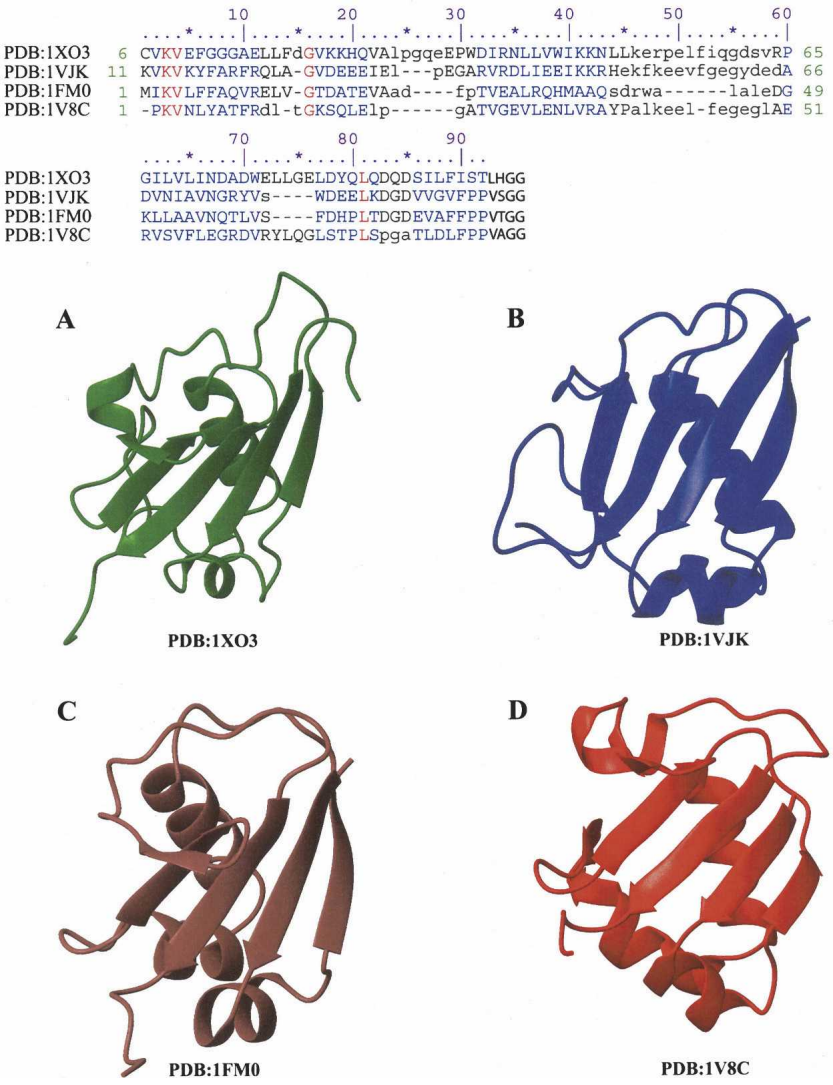

Figure 5. Sequence alignment (*top*) and structural comparison (*bottom*) of the structural homologs. (*A*) AAH26994.1 from *Mus musculus* (1XO3); (*B*) molybdopterin converting factor from *Pyrococcus furiosus* (1VJK); (*C*) molybdopterin synthase from *Escherichia coli* (1FM0); (*D*) MoaD related protein from *Thermus thermophilus* (1V8C).

respectively (Fig. 5). For comparison, AAH26994.1 and human ubiquitin have a 15% sequence identity over 59 residues and a C^{α} RMSD of 3.5 Å (*Z*-score of 4.2) for the corresponding structures.

Our current understanding of the putative target(s) and function(s) of Urm1 pathway is sparse. Although Urm1 was found not essential for growth, it showed a temperature-sensitivity phenotype of growth retardation (Furukawa et al. 2000). More recent work has shown that Urm can modify a substrate protein: Urm1 has been found to conjugate with alkyl hydroperoxide reductase (Ahp1), a protein implicated in oxidative-stress reduction in yeast (Goehring et al. 2003a). In addition, the Urm1 pathway has been shown to be involved in nutrient sensing, invasive growth, and budding in yeast (Goehring et al. 2003b). Thus Urm1 can be conjugated to a variety of proteins, and the conjugation pathway appears to play roles in both vegetative and invasive growth.

Furukawa et al. (2000) speculated that the Urm1 conjugation system "may provide a missing link between ATP-dependent cofactor sulfuration and ATP-dependent protein conjugation." The striking structural similarity between the structure of the ubiquitin-like protein Urm1 reported here and the structures of bacterial MoaD proteins (Fig. 5) greatly strengthens this hypothesis. All three closely similar structures are from members of a class of proteins called molybdopterin converting factor/molybdopterin synthase. This protein acts as a sulfur donor in molybdenum cofactor (Moco) biosynthesis (Rudolph et al. 2001, 2003), a component of

a biosynthetic pathway present in eubacteria, archaea, and eukaryotes, including humans. These structural results lend considerable support to the hypothesis (Furukawa et al. 2000) that the ubiquitin and ubiquitin-like conjugation systems found in eukaryotes evolved from primitive sulfur transfer systems in more primitive organisms.

Materials and methods

CESG's laboratory information management system, Sesame (Zolnai et al. 2003), was used to track all operations in this project. Unless otherwise stated, bacterial growth reagents, antibiotics, routine laboratory chemicals, and disposable labware were from Sigma-Aldrich, Fisher, or other major distributors. The 2-L polyethylene terephthalate (PET) beverage bottles used for bacterial cell growth were from Ball Corp.

Cloning, expression, and protein purification

The mouse ORF clone of AAH26994 was obtained from the IMAGE consortium's Mammalian Gene Collection (Strausberg et al. 1999). The ORF was reamplified and subcloned according to the protocols described in Thao et al. (2005). The expression vector used was our custom-made, pQE80-based (Qiagen) expression vector pVP16-GW. This Gateway (Invitrogen)-derived vector creates a fusion of the maltose-binding protein, 8XHIS tag, and tobacco etch virus (TEV) protease cleavage site to the N terminus of the mouse ORF.

Escherichia coli B834 (Novagen) was used as the expression host for isotopic labeling. This strain likely has a mutation in *metE*, and thus requires either methionine or vitamin B_{12} for growth. The cells were made competent by the Z-Competent (Geno Technology) method and then transformed with pLacI RARE (Novagen). The pLacI RARE transformants were also made competent by the same method.

Expression of double-labeled AAH26994 was carried out in 2 L of chemically defined auto-induction medium. The auto-induction medium was a modification of that originally developed by F. William Studier at Brookhaven National Laboratory (Studier 2005). The auto-induction growth components were modified to include isotopically labeled compounds, [*U*-^{15}N]-labeled NH_4Cl, [*U*-^{13}C]-glucose, and [*U*-^{13}C]-glycerol as the nitrogen and carbon sources, which allowed for [*U*-^{13}C;*U*-^{15}N] labeling of target protein. The exact details of the procedure, as well as the modified medium components for ^{13}C;^{15}N labeling using autoinduction have been described elsewhere (Tyler et al. 2005). Following this method, 18.7 g of wet cell mass were produced from 2 L of growth medium.

The cell mass was suspended in 75 mL of 50 mM NaH_2PO_4 (pH 7.5), containing 0.5 M NaCl, 20% (w/v) ethylene glycol and 35 mM imidazole (Jeon et al. 2005). The cells were lysed by sonication. The sonicated cells were clarified by centrifugation. The recombinant fusion protein was purified by Ni^{2+}-IMAC chromatography in a linear 0–500-mM imidazole gradient. The purified fusion protein was desalted into 20 mM NaH_2PO_4 (pH 7.5), containing 100 mM NaCl, and reacted with tobacco etch virus (TEV) protease overnight at 25°C. The proteolyzed target was subjected to subtractive Ni^{2+}-IMAC chromatography, and the pure target was desalted into 5 mM HEPES (pH 7.0), containing 50 mM NaCl, and concentrated. For final NMR experiments, the protein was exchanged with 10 mM Bis-Tris buffer containing 50 mM NaCl (pH 6.5).

NMR measurements

NMR samples consisted of 1.0 mM [*U*-^{13}C;*U*-^{15}N]-AAH26994 in 5 mM Bis-Tris (pH 6.5). NMR spectra were collected at 25°C on a Varian ^{UNITY}INOVA 800 MHz NMR spectrometer equipped with a 5-mm *z*-shielded gradient ^{1}H-^{13}C-^{15}N triple-resonance cold probe. For the backbone resonance assignments, three-dimensional (3D) HNCO, HNCACB, CBCA(CO)NH, NOESY-(^{15}N,^{1}H)-HSQC spectra were recorded as described in Palmer et al. (1992), and the side-chain resonance assignments, 3D HBHA(CO)NH, H(CCO)NH, C(CO)NH, and H(C)CH TOCSY data sets were recorded. NOE distance constraints were obtained from 3D-NOESY(^{15}N,^{1}H)-HSQC, and 3D NOESY-(^{13}C,^{1}H)-HSQC spectra with 100-msec mixing times. The software package NMRPipe (Delaglio et al. 1995) was used to process all NMR time-domain data, and NMRView (Johnson and Blevins 1994) was used to analyze spectra.

Chemical shift-derived and hydrogen-bond restraints

Assigned chemical shifts were determined for 97% of the nuclei. The $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, $^{1}H^{\alpha}$, and ^{15}N chemical shifts of the assigned residues served as input for the TALOS program (Cornilescu et al. 1999). TALOS derives information on the ϕ and ψ backbone dihedral angles from a comparison of secondary chemical-shift patterns of amino acid triplets against a database of secondary chemical shifts corresponding to known conformations. A conservative approach was chosen requiring that all 10 best matches agree for a prediction to be accepted. The TALOS predictions were converted into dihedral angle restraints as the average ϕ and ψ angles $\pm 2\times$ the standard deviation or a minimum of $\pm 20°$. For better convergence, a number of hydrogen-bond restraints were introduced for the backbone amide protons on the basis of amide ^{1}H-^{2}H exchange results, $^{13}C^{\alpha}/^{13}C^{\beta}$ secondary shifts, and NOE data. Hydrogen bonds were enforced by using the following restraints: 1.8–2.3 Å for $d_{(N\text{-}H,\ O)}$; 2.7–3.3 Å for $d_{(N,\ O)}$. The hydrogen bonds between N-H_i and O-C_j, in the β-sheet structures were included as restraints only if the β-sheet interstrand $d_{NN}(i, j)$ and $d_{\alpha N}(i, j)$ NOE cross-peaks were observed. Hydrogen-bond constraints for α-helices were included when NOEs corresponding to the secondary structure $d_{\alpha\ N}(i, i+3)$ for α-helices were observed.

Structure calculations

All calculations were performed with CNS (Brünger et al. 1998) using the ARIA setup and protocols (Nilges and O'Donoghue 1998; Linge and Nilges 1999). The TALOS-derived dihedral angles were restrained with a harmonic potential using a force constant of 200 kcal mol^{-1} rad^{-2}. Covalent interactions were calculated with a modified version of the PARALLHDG 5.3 parameter file (Linge and Nilges 1999) based on the CSDX parameter set. In addition to the bonded energy terms typically used in NMR structure calculations (bond, angle, and improper energy terms), the dihedral angle energy term describing torsions around rotatable bonds ("dihe" flag in CNS) was turned on. This energy term greatly improved the quality of side-chain χ^1 and χ^2 rotamers as assessed by PROCHECK (Laskowski et al. 1996). Nonbond-

ed interactions were calculated with the repel function by using the PROLSQ parameters as implemented in the PARALLHDG parameter file. The OPLS nonbonded parameters (Jorgensen 1988) were used for the final water refinement, including full van der Waals and electrostatic energy terms. The nonbonded pair list was generated with a 9.5 Å cutoff, and the nonbonded interaction was calculated with an 8.5 Å cutoff using a shifting function.

A simulated annealing protocol in Cartesian space was used, starting from an extended conformation. Force constants were scaled throughout the protocol following the default ARIA/CNS setup. The atomic masses were set uniformly to 100 amu, and the friction coefficient f_{beta} for the coupling to the external temperature bath was set to 20 $psec^{-1}$. The simulated annealing protocol, which is similar to the one described in Linge and Nilges (1999), consisted of four stages: (1) high-temperature SA stage (10,000 steps; 2000 K), (2) a first cooling phase from 2000 to 1000 K in 10,000 steps, (3) a second cooling phase from 1000 K to 50 K in 5000 steps, followed by (4) 200 steps of energy minimization. The time step for the integration was set to 0.003 psec.

The structures were subjected to a final refinement protocol in explicit water by solvating them with a 8-Å layer of TIP3P waters (Duffy et al. 1992). The water refinement consisted of a heating period (50 MD steps at 100, 200, 300, 400, and 500 K; time step 0.005 fsec) with harmonic position restraints on the C^{α} atoms (kharm = 10 kcal $mol^{-1}Å^{2}$), followed by 2500 MD steps at 500 K without any position restraints, and a final cooling stage from 500 to 100 K in 100 K steps (1000 MD steps per temperature step). The resulting structures were energy minimized with 100 steps of Powell steepest descent minimization.

PROCHECK-NMR (Laskowski et al. 1996) was used to assess the quality of the final ensemble of conformers. Structures were visualized with the programs MOLMOL (Koradi et al. 1996) or RASMOL (Sayle and Milner-White 1995). The chemical shifts of AAH26994.1 at pH 6.5 and 298 K have been deposited to BioMagResBank under accession number 6337. The atomic coordinates of the ensemble of 20 structures that represent the solution structure of AAH26994.1 have been deposited in the PDB, together with the complete list of restraints used for structure calculation under accession number 1XO3. The structure was determined under the NIH, NIGMS Protein Structure Initiative.

Acknowledgments

We thank Hamid Eghbalnia for advice in using the PISTACHIO automated assignment software and other members of the CESG project whose contributions made this work possible. We are most grateful to two reviewers who first pointed out that the structure reported here most probably is that of a mouse Urm1. This study was supported by the NIH, Protein Structure Initiative through grant P50 GM64598; NMR data were collected at the National Magnetic Resonance Facility at Madison, which is supported in part by grants P41 RR02301 and P41 GM66326.

References

Brünger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta. Crystallogr. D Biol. Crystallogr.* **54:** 905–921.

Cornilescu, G., Delaglio, F., and Bax, A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13:** 289–302.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6:** 277–293.

Duffy, E.M., Severance, D.L., and Jorgensen, W.L. 1992. Solvent effects on the barrier to isomerization for a tertiary amide from Ab initio and Monte-Carlo calculations. *J. Am. Chem. Soc.* **114:** 7535–7542.

Eghbalnia, H., Bahrami, A., Wang, L., Assadi, A., and Markley, J.L. 2005. PISTACHIO: An approach for probabilistic sequence-specific assignments of signals from protein NMR spectra. *J. Biomol. NMR* (in press).

Furukawa, K., Mizushima, N., Noda, T., and Ohsumi, Y. 2000. A protein conjugation system in yeast with homology to biosynthetic enzyme reaction of prokaryotes. *J. Biol. Chem.* **275:** 7462–7465.

Gibrat, J.F. and Madej, T.B.S. 1996. Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.* **6:** 377–385.

Goehring, A.S., Rivers, D.M., and Sprague Jr., G.F. 2003a. Urmylation: A ubiquitin-like pathway that functions during invasive growth and budding in yeast. *Mol. Biol. Cell* **14:** 4329–4341.

———. 2003b. Attachment of the ubiquitin-related protein Urm1p to the antioxidant protein Ahp1p. *Eukaryotic Cell* **2:** 930–936.

Hershko, A., Ciechanover, J.A., and Varshavsky, A. 2000. The ubiquitin system. *Nat. Med.* **6:** 1073–1081.

Hochstrasser, M. 2000. Evolution and function of ubiquitin-like protein-conjugation systems. *Nat. Cell Biol.* **2:** E153–E157.

Holm, L. and Sander, C. 1995. Dali: A network tool for protein structure comparison. *Trends. Biochem. Sci.* **20:** 478–480.

Jeon, W.B., Aceti, D.J., Bingman, C.A., Vojtik, F.C., Olson, A.C., Ellefson, J.M., McCombs, J.E., Sreenath, H., Blommel, P.G., Seder, K.D., et al. 2005. High-throughput purification and quality assurance of *Arabidopsis thaliana* proteins for eukaryotic structural genomics. *J. Struct. Funct. Genomics* (in press).

Johnson, B.A. and Blevins, R.A. 1994. NMRView–A computer-program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4:** 603–614.

Jorgensen, W.L. 1988. OPLS force fields. *Encycl. Computat. Chem.* **3:** 1754–1763.

Koradi, R., Billeter, M., and Wüthrich, K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14:** 51–55.

Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. 1996. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8:** 477–486.

Linge, J.P. and Nilges, M. 1999. Influence of non-bonded parameters on the quality of NMR structures: A new force field for NMR structure calculation. *J. Biomol. NMR* **13:** 51–59.

Nilges, M. and O'Donoghue, S.I. 1998. Ambiguous NOEs and automated NOE assignment. *Prog. NMR Spect.* **32:** 107–139.

Palmer III, A.G., Fairbrother, W.J., Cavanagh, J., Wright, P.E., and Rance, M. 1992. Improved resolution in three-dimensional constant-time triple resonance NMR spectroscopy of proteins. *J. Biomol. NMR* **2:** 103–108.

Rudolph, M.J., Wuebbens, M.M., Rajagopalan, K.V., and Schindelin, H. 2001. Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation. *Nat. Struct. Biol.* **8:** 42–46.

Rudolph, M.J., Wuebbens, M.M., Turque, O., Rajagopalan, K.V., and Schindelin, H. 2003. Structural studies of molybdopterin synthase provide insights into its catalytic mechanism. *J. Biol. Chem.* **278:** 14514–14522.

Sayle, R.A. and Milner-White, E.J. 1995. RASMOL: Biomolecular graphics for all. *Trends Biochem. Sci.* **20:** 374.

Strausberg, R.L., Feingold, E.A., Klausner, R.D., and Collins, F.S. 1999. The mammalian gene collection. *Science* **286:** 455–457.

Studier, F.W. 2005. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41:** 207–234.

Thao, S., Zhao, Q., Kimball, T., Steffen, E., Blommel, P.G., Riters, M., Newman, C.S., Fox, B.G., and Wrobel, R.L. 2005. Results from high-throughput DNA cloning of *Arabidopsis thaliana* target genes using site-specific recombination. *J. Struct. Funct. Genomics* **5:** 267–276.

Tyler, R.C., Sreenath, H., Aceti, D.J., Bingman, C.A., Singh, S., Markley, J.L., and Fox, B.G. 2005. Auto-induction medium for the production of [*U*-^{15}N]- and [*U*-^{13}C, *U*-^{15}N]-labeled proteins for NMR screening and structure determination. *Protein Expr. Purif.* **40:** 268–278.

Wang, C., Xi, J., Begley, T.P., and Nicholson, L.K. 2001. Solution structure of ThiS and implications for the evolutionary roots of ubiquitin. *Nat. Struct. Biol.* **8:** 47–51.

Wishart, D.S. and Sykes, B.D. 1994. The ^{13}C chemical shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical shifts. *J. Biomol. NMR* **4:** 171–180.

Zolnai, Z., Lee, P.T., Li, J., Chapman, M.R., Newman, C.S., Phillips Jr., G.N., Rayment, I., Ulrich, E.L., Volkman, B.F., and Markley, J.L. 2003. Project management system for structural and functional proteomics: Sesame. *J. Struct. Funct. Genomics* **4:** 11–23.